

SEQUENCE LISTING

<110> Donnelly, John J.
 Liu, Margaret A.
 Shiver, John W.
 Fu, Tong-Ming

<120> SYNTHETIC HEPATITIS C GENES

<130> 19732YPCA

<140> 09/194,949

<141> 2000-02-17

<150> PCT/US97/09884

<151> 1997-06-06

<150> 60/033,534

<151> 1996-12-20

<150> 60/020,494

<151> 1996-06-11

<160> 25

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 3610

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 1

```

gatattggct attggccatt gcatacgttg tatccatatac ataatatgta catttatatt 60
ggctcatgtc caacattacc gccatgttga cattgattat tgactagtta ttaatagtaa 120
tcaattacgg ggctcattagt tcatagccca tatatggagt tccgcgttac ataacttacg 180
gtaaatggcc cgcctggctg accgccaac gacccccgcc cattgacgtc aataatgacg 240
tatgttccca tagtaacgcc aatagggact ttccattgac gtcaatgggt ggagtattta 300
cggtaaactg cccacttggc agtacatcaa gtgtatcata tgccaagtac gccccctatt 360
gacgtcaatg acggtaaatg gcccgcttg cattatgccc agtacatgac cttatgggac 420
tttcctactt ggcagtacat ctacgtatta gtcacgcta ttaccatggg gatgcggttt 480
tggcagtaca tcaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccac 540
cccattgacg tcaatgggag tttgttttgg caccaaaatc aacgggactt tccaaaatgt 600
cgtaacaact ccgccccatt gacgcaaatg ggcggtaggc gtgtacgggt ggaggtctat 660
ataagcagag ctcgtttagt gaaccgtcag atcgccctgga gacgccatcc acgctgtttt 720
gacctccata gaagacaccg ggaccgatcc agcctccgcg gccgggaacg gtgcattgga 780
acgcggattc cccgtgccaa gagtgcgta agtaccgcct atagagtcta tagggcccacc 840
cccttggtct cttatgcatg ctatactgtt tttggcttgg ggtctataca cccccgttc 900
ctcatgttat aggtgatggt atagcttagc ctataggtgt gggttattga ccattattga 960
ccactcccct attggtgacg atactttcca ttactaatcc ataacatggc tctttgccac 1020
aactctcttt attggctata tgccaatata ctgtccttca gagactgaca cggactctgt 1080
atttttacag gatgggggtct catttattat ttacaaattc acatatacaa caccaccgtc 1140

```

```

cccagtgccc gcagttttta ttaaacataa cgtgggatct ccacgcgaat ctcggttacg 1200
tggtccggac atggggtctt ctccggtagc ggcggagctt ctacatccga gccctgctcc 1260
catgcctcca gcgactcatg gtcgctcggc agctccttgc tcctaacagt ggaggccaga 1320
cttaggcaca gcacgatgcc caccaccacc agtgtgccgc acaaggccgt ggcggtaggg 1380
tatgtgtctg aaaatgagct cggggagcgg gcttgaccgc ctgacgcatt tggaagactt 1440
aaggcagcgg cagaagaaga tgcaggcagc tgagttgttg tggtctgata agagtcagag 1500
gtaactcccg ttgcggtgct gttaacgggt gagggcagtg tagtctgagc agtactcgtt 1560
gctgccgcgc gcgccaccag acataatagc tgacagacta acagactgtt cctttccatg 1620
ggtcttttct gcagtcaccg tccttagatc taggtaccag atatcagaat tcagtcgaca 1680
gcggccgcga tctgctgtgc ctctagttg ccagccatct gttgtttgcc cctccccgt 1740
gccttctctg accctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat 1800
tgcatacgcat tgtctgagta ggtgtcattc tattctgggg ggtggggtgg ggcagcacag 1860
caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg gctctatggg 1920
tacggccgca gcggccttaa ttaaggccgc agcggccgta ccaggtgct gaagaattga 1980
cccgttctct cgacccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc 2040
cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 2100
taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgacctg 2160
ccgcttaccg gatacctgtc cgcttttctc ccttcgggaa gcgtggcgct ttctcaatgc 2220
tcacgctgta ggtatctcag ttcggtgtag gtcttctgct ccaagctggg ctgtgtgcac 2280
gaaccccccg ttcagcccga ccgctgcgcc ttatccggtg actatcgtct tgagtccaac 2340
ccggttaagac acgacttatt gccactggca gcagccactg gtaacaggat tagcagagcg 2400
aggatgtag gcggtgctac agagtctctg aagtgggtggc ctaactacgg ctacactaga 2460
aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaag aagagttggt 2520
agctcttgat ccggcaaaaa aaccaccgct ggtagcgggt gtttttttgt ttgcaagcag 2580
cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgtgatcc 2640
cgtaatgctc tgccagtgtt acaaccaatt aaccaattct gattagaaaa actcatcgag 2700
catcaaatga aactgcaatt tattcatatc aggattatca ataccatatt tttgaaaaag 2760
ccgtttctgt aatgaaggag aaaactcacc gaggcagttc cataggatgg caagatcctg 2820
gtatcggtct gcgattccga ctctccaac atcaatacaa cctattaatt tcccctcgtc 2880
aaaaataagg ttatcaagtg agaaatcacc atgagtgcgc actgaatccg gtgagaatgg 2940
caaaagctta tgcatttctt tccagacttg ttcaacaggc cagccattac gctcgtcatc 3000
aaaatcactc gcatcaacca aaccgttatt cattcgtgat tgcgcctgag cgagacgaaa 3060
tacgcgatcg ctgttaaaaag gacaattaca aacaggaatc gaatgcaacc ggcgaggaa 3120
cactgccagc gcatcaacaa tattttcacc tgaatcagga tattcttcta atacctggaa 3180
tgctgttttc ccggggatcg cagtgggtgag taacctatgca tcacaggag tacggataaa 3240
atgcttgatg gtcggaagag gcataaattc cgtcagccag tttagtctga ccattctcatc 3300
tgtaacatca ttggcaacgc tacctttgcc atgtttcaga aacaactctg gcgcatcggg 3360
cttcccatat aatcgataga ttgtcgacc tgattgcccg acattatcgc gagccattt 3420
atacccatat aaatcagcat ccattgttga atttaatcgc ggcctcgagc aagacgtttc 3480
ccgttgaata tggctcataa cacccttgtt attactgtt atgtaagcag acagttttat 3540
tgttcatgat gatataatct tatcttgtgc aatgtaacat cagagatttt gagacacaac 3600
gtggctttcc 3610

```

<210> 2

<211> 573

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized sequence encoding HCV core antigen

<400> 2

```

atgagcacca accccaagcc ccagaggaag accaagagga acaccaacag gagggccccag 60
gatgtgaagt tccctggggg aggccagatt gtgggagggg tctacctgct gcccaggagg 120
ggccccaggc tgggggtgag ggctaccagg aagacctctg agagggtccca gcccaggggc 180

```

```

aggaggcagc ccatacccaa ggccaggagg cctgagggcc gctcctgggc ccagcctggc 240
tacccttggc ccctgtatgg caatgaaggc ttggctggg ctggctggct gctgtcccc 300
aggggctcca ggccctcctg gggcccccaca gacccagga ggaggtccag gaacctgggc 360
aaggtgattg acaccctgac ctgtggcctt gctgacctga tgggctacat cccctgggtg 420
ggggctcctg tgggaggggt ggctagggct ctggctcatg gggtgagggg gctggaggat 480
ggggtgaact atgctactgg caacctgcct ggctgctcct tctccatctt cctgctggcc 540
ctgctctcct gcctgacagt gcctgcttct gcc 573

```

<210> 3
 <211> 191
 <212> PRT
 <213> Hepatitis C Virus

```

<400> 3
Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1          5          10          15
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 20          25          30
Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
 35          40          45
Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
 50          55          60
Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Ser Trp Ala Gln Pro Gly
 65          70          75          80
Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Phe Gly Trp Ala Gly Trp
 85          90          95
Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
100          105          110
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
115          120          125
Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Val
130          135          140
Gly Gly Val Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145          150          155          160
Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile
165          170          175
Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala
180          185          190

```

<210> 4
 <211> 103
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified Vector Sequence

```

<400> 4
gaattcgctt ccaatgagaa catggagacc atgaaccagc cctaccacat ctgccgcggc 60
ttcacctgct tcaagaagta aaccgggaa ttctaaagtc gac 103

```

<210> 5
 <211> 573
 <212> DNA
 <213> Hepatitis C Virus

```

<400> 5
atgagcacga atcctaaacc tcaaagaaaa accaaacgta acaccaaccg ccgcccacag 60
gacgtcaagt tcccgggagg tggtcagatc gttgggtggag tttacttggt gccgcgcagg 120
ggccccaggt tgggtgtgag cgcgactagg aagacttccg agcggtcgca acctcgtgga 180
aggcgacagc ctatcccca aagctcgccg cccgagggca ggtcctgggc tcagccccgg 240
tacccttggc ccctctatgg caatgagggc ttcgggtggg caggatggct cctgtcccc 300
cgcggctctc ggcctagttg gggccccact gacccccggc gtaggtcgcg caatttgggt 360
aaggtcatcg ataccctcac gtgcggcttc gccgacctca tggggtacat cccgctcgtc 420
ggcgcccccg tagggggcgt cgccaggggc ctggcgcatg gcgtcagggt tctggaggac 480
ggggtgaact atgcaacagg gaatttgccc ggttgctctt tctctatctt cctcctgggt 540
ctgctgtcct gcctgaccgt cccagcttct gct 573

```

<210> 6

<211> 582

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized sequence encoding HCV E1 protein

```

<400> 6
atgtatgagg tgaggaatgt ctctggcgctc taccatgtga ccaatgactg ctccaactcc 60
tgcatgtgct atgaggctgc tgacatgac atgcacaccc ctggctgtgt gccatgtgtg 120
agggagggca actcctccag gtgctgggtg gccctgaccc ccaccctggc tgccaggaac 180
tcttccatcc ccaccaccac catcaggagg catgtggacc tgctgggtggg cgctgctgcc 240
ctgtgctctg ccatgtatgt gggcgacctg tgtggctctg tcttctggt gtcccagctg 300
ttcaccttct ccccaggag gtatgagact gtgcaggact gcaactgctc cctgtaccct 360
ggccatgtct ctggccacag gatggcctgg gacatgatga tgaactggtc cccaccact 420
gccctgggtg tctcccagct gctgaggatc ccccaggctg tgggtggacat ggtgggtggg 480
gcccactggg gcgtgctggc tggcctggcc tactactcca tgggtgggcaa ctggggccaa 540
gtgctgattg tgatgctgct gtttgctggc gtggatggct aa 582

```

<210> 7

<211> 193

<212> PRT

<213> Hepatitis C Virus

<400> 7

```

Met Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp
1           5           10           15
Cys Ser Asn Ser Cys Ile Val Tyr Glu Ala Ala Asp Met Ile Met His
20           25           30
Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser Arg Cys
35           40           45
Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser Ser Ile Pro
50           55           60
Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala
65           70           75           80
Leu Cys Ser Ala Met Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu
85           90           95
Val Ser Gln Leu Phe Thr Phe Ser Pro Arg Arg Tyr Glu Thr Val Gln
100          105          110
Asp Cys Asn Cys Ser Leu Tyr Pro Gly His Val Ser Gly His Arg Met
115          120          125

```

```

Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu Val Val
 130                      135                      140
Ser Gln Leu Leu Arg Ile Pro Gln Ala Val Val Asp Met Val Val Gly
 145                      150                      155                      160
Ala His Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly
                      165                      170                      175
Asn Trp Ala Lys Val Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp
                      180                      185                      190
Gly

```

<210> 8
 <211> 1044
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Optimized sequence encoding HCV E2 protein

```

<400> 8
atgaccacct atgtctctgt gggccatgcc tcccagacca ccaggagggt ggccctccttc 60
ttctccccctg gctctgcccc gaagatccag ctggtgaaca ccaatggctc ctggcacatc 120
aacaggactg ccctgaattg caacgagtc atcaacactg gcttctttgc tgccctgttc 180
tatgtgaaga agttcaactc ctctggctgc tctgagagga tggcctcctg caggcccatt 240
gacaggtttg cccagggctg gggcccatc acccatgctg agtccaggtc ctctgaccag 300
aggccatact gctggcacta tgccccccag ccatgtggca ttgtgectgc cctgcaggtc 360
tgtggccctg tctactgctt caccaccatc cctgtggtgg tgggcaccac tgacaggttt 420
ggcgtgcccc cctacaactg gggcgacaat gagactgatg tgctgctgct gaacaacacc 480
aggccccccc agggcaactg gtttggtgc acctggatga actccactgg cttcaccaag 540
acctgtggcg gcccccatg caacattggc ggcgctggca acaacaccct gacctgcccc 600
actgactgct tcaggaagca tcctgaggcc acctacacca agtgtggctc tggcccatgg 660
ctgaccccca ggtgcatggt ggactaccca tacaggctgt ggcactaccc atgcaccttc 720
aacttcacca tcttcaagat caggatgtat gtgggcggcg tggagcacag gctgaatgct 780
gcctgcaact ggaccagggg cgagagggtgc aacattgagg acagggacag gtctgagctg 840
tccccctgct tgctgtccac cactgagtgg cagatcctgc catgctcctt caccacctgc 900
cctgcctgtt ccactggcct gatccatctg catcagaaca ttgtggatgt gcagtacctg 960
tacggcgtgg gctccgctgt ggtctccatt gtgatcaagt gggagtatgt gctgctgctg 1020
ttcctgctgc tggctgatgc cttaa 1044

```

<210> 9
 <211> 347
 <212> PRT
 <213> Hepatitis C Virus

```

<400> 9
Met Thr Thr Tyr Val Ser Val Gly His Ala Ser Gln Thr Thr Arg Arg
 1                      5                      10                      15
Val Ala Ser Phe Phe Ser Pro Gly Ser Ala Gln Lys Ile Gln Leu Val
                      20                      25                      30
Asn Thr Asn Gly Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn
                      35                      40                      45
Glu Ser Ile Asn Thr Gly Phe Phe Ala Ala Leu Phe Tyr Val Lys Lys
                      50                      55                      60
Phe Asn Ser Ser Gly Cys Ser Glu Arg Met Ala Ser Cys Arg Pro Ile
                      65                      70                      75                      80

```

Asp	Arg	Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	Thr	His	Ala	Glu	Ser	Arg	
				85					90					95		
Ser	Ser	Asp	Gln	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Ala	Pro	Gln	Pro	Cys	
			100					105					110			
Gly	Ile	Val	Pro	Ala	Leu	Gln	Val	Cys	Gly	Pro	Val	Tyr	Cys	Phe	Thr	
		115					120					125				
Pro	Ser	Pro	Val	Val	Val	Gly	Thr	Thr	Asp	Arg	Phe	Gly	Val	Pro	Thr	
		130				135					140					
Tyr	Asn	Trp	Gly	Asp	Asn	Glu	Thr	Asp	Val	Leu	Leu	Leu	Asn	Asn	Thr	
145					150					155					160	
Arg	Pro	Pro	Gln	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met	Asn	Ser	Thr	
			165					170						175		
Gly	Phe	Thr	Lys	Thr	Cys	Gly	Gly	Pro	Pro	Cys	Asn	Ile	Gly	Gly	Ala	
			180					185					190			
Gly	Asn	Asn	Thr	Leu	Thr	Cys	Pro	Thr	Asp	Cys	Phe	Arg	Lys	His	Pro	
		195					200					205				
Glu	Ala	Thr	Tyr	Thr	Lys	Cys	Gly	Ser	Gly	Pro	Trp	Leu	Thr	Pro	Arg	
	210					215					220					
Cys	Met	Val	Asp	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	Thr	Phe	
225					230					235					240	
Asn	Phe	Thr	Ile	Phe	Lys	Ile	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	
			245					250						255		
Arg	Leu	Asn	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asn	Ile	
		260						265					270			
Glu	Asp	Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	Thr	Thr	
	275						280					285				
Glu	Trp	Gln	Ile	Leu	Pro	Cys	Ser	Phe	Thr	Thr	Leu	Pro	Ala	Leu	Ser	
	290					295					300					
Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	Tyr	Leu	
305					310					315					320	
Tyr	Gly	Val	Gly	Ser	Ala	Val	Val	Ser	Ile	Val	Ile	Lys	Trp	Glu	Tyr	
			325					330						335		
Val	Leu	Leu	Leu	Phe	Leu	Leu	Leu	Ala	Asp	Ala						
			340					345								

<210> 10

<211> 1620

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized sequence encoding HCV E1 + E2 proteins

<400> 10

atgtatgagg	tgaggaatgt	ctctggcgtc	taccatgtga	ccaatgactg	ctccaactcc	60
tgcatgtgt	atgaggctgc	tgacatgatc	atgcacaccc	ctggctgtgt	gccatgtgtg	120
agggagggca	actcctccag	gtgctgggtg	gccctgaccc	ccaccctggc	tgccaggaac	180
tcctccatcc	ccaccaccac	catcaggagg	catgtggacc	tgctggtggg	cgctgctgcc	240
ctgtgctctg	ccatgtatgt	ggcgacactg	tgtggctctg	tcttctgggt	gtcccagctg	300
ttcaccttct	cccccaggag	gtatgagact	gtgcaggact	gcaactgctc	cctgtaccct	360
ggccatgtct	ctggccacag	gatggcctgg	gacatgatga	tgaactggtc	ccccaccact	420
gccctggtgg	tctcccagct	gctgaggatc	ccccaggctg	tggtggacat	ggtggtgggc	480
gcccactggg	gcgtgctggc	tggcctggcc	tactactcca	tggtgggcaa	ctgggccaag	540
gtgctgattg	tgatgctgct	gttgctggc	gtggatggca	ccacctatgt	ctctgtgggc	600
catgcctccc	agaccaccag	gaggggtggc	tccttcttct	cccctggctc	tgcccagaag	660

```

atccagctgg tgaacaccaa tggctcctgg cacatcaaca ggactgccct gaattgcaac 720
gagtccatca acactggctt ctttgctgcc ctgttctatg tgaagaagtt caactcctct 780
ggctgctctg agaggatggc ctccctgcagg cccattgaca ggtttgccca gggctggggc 840
cccatcacc c atgctgagtc caggctcctct gaccagaggc catactgctg gcactatgcc 900
ccccagccat gtggcattgt gcctgccctg caggctctgtg gccctgtcta ctgcttcacc 960
ccatcccctg tgggtgggtggg caccactgac aggtttggcg tgcccaccta caactggggc 1020
gacaatgaga ctgatgtgct gctgctgaac aacaccaggc cccccaggg caactggttt 1080
ggctgcacct ggatgaactc cactggcttc accaagacct gtggcggccc cccatgcaac 1140
attggcggcg ctggcaacaa caccctgacc tgcccactg actgcttcag gaagcatcct 1200
gagggccact acaccaagtg tggctctggc ccatggctga cccccagggtg catggtggac 1260
taccatacaga ggctgtggca ctacccatgc accttcaact tcaccatctt caagatcagg 1320
atgtatgtgg gcggcgtgga gcacaggctg aatgctgcct gcaactggac caggggctgag 1380
aggtgcaaca ttgaggacag ggacaggctc gagctgtccc ccctgctgct gtccaccact 1440
gagtggcaga tcctgccatg ctcccttcacc accctgcctg ccctgtccac tggcctgatc 1500
catctgcata agaacattgt ggatgtgcag tacctgtacg gcgtgggctc cgctgtgggtc 1560
tccattgtga tcaagtggga gtatgtgctg ctgctgttcc tgctgctggc tgatgcctaa 1620

```

<210> 11

<211> 539

<212> PRT

<213> Hepatitis C Virus

<400> 11

```

Met Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp
 1          5          10          15
Cys Ser Asn Ser Cys Ile Val Tyr Glu Ala Ala Asp Met Ile Met His
          20          25          30
Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser Arg Cys
          35          40          45
Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser Ser Ile Pro
          50          55          60
Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala
          65          70          75          80
Leu Cys Ser Ala Met Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu
          85          90          95
Val Ser Gln Leu Phe Thr Phe Ser Pro Arg Arg Tyr Glu Thr Val Gln
          100          105          110
Asp Cys Asn Cys Ser Leu Tyr Pro Gly His Val Ser Gly His Arg Met
          115          120          125
Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu Val Val
          130          135          140
Ser Gln Leu Leu Arg Ile Pro Gln Ala Val Val Asp Met Val Val Gly
          145          150          155          160
Ala His Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly
          165          170          175
Asn Trp Ala Lys Val Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp
          180          185          190
Gly Thr Thr Tyr Val Ser Val Gly His Ala Ser Gln Thr Thr Arg Arg
          195          200          205
Val Ala Ser Phe Phe Ser Pro Gly Ser Ala Gln Lys Ile Gln Leu Val
          210          215          220
Asn Thr Asn Gly Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn
          225          230          235          240
Glu Ser Ile Asn Thr Gly Phe Phe Ala Ala Leu Phe Tyr Val Lys Lys
          245          250          255

```

```

Phe Asn Ser Ser Gly Cys Ser Glu Arg Met Ala Ser Cys Arg Pro Ile
      260      265      270
Asp Arg Phe Ala Gln Gly Trp Gly Pro Ile Thr His Ala Glu Ser Arg
      275      280      285
Ser Ser Asp Gln Arg Pro Tyr Cys Trp His Tyr Ala Pro Gln Pro Cys
      290      295      300
Gly Ile Val Pro Ala Leu Gln Val Cys Gly Pro Val Tyr Cys Phe Thr
      305      310      315      320
Pro Ser Pro Val Val Gly Thr Thr Asp Arg Phe Gly Val Pro Thr
      325      330      335
Tyr Asn Trp Gly Asp Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr
      340      345      350
Arg Pro Pro Gln Gly Asn Trp Phe Gly Cys Thr Trp Met Asn Ser Thr
      355      360      365
Gly Phe Thr Lys Thr Cys Gly Gly Pro Pro Cys Asn Ile Gly Gly Ala
      370      375      380
Gly Asn Asn Thr Leu Thr Cys Pro Thr Asp Cys Phe Arg Lys His Pro
      385      390      395      400
Glu Ala Thr Tyr Thr Lys Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg
      405      410      415
Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Phe
      420      425      430
Asn Phe Thr Ile Phe Lys Ile Arg Met Tyr Val Gly Gly Val Glu His
      435      440      445
Arg Leu Asn Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys Asn Ile
      450      455      460
Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Ser Thr Thr
      465      470      475      480
Glu Trp Gln Ile Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser
      485      490      495
Thr Gly Leu Ile His Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu
      500      505      510
Tyr Gly Val Gly Ser Ala Val Val Ser Ile Val Ile Lys Trp Glu Tyr
      515      520      525
Val Leu Leu Leu Phe Leu Leu Leu Ala Asp Ala
      530      535

```

<210> 12

<211> 1350

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized sequence encoding HCV NS5a protein

<400> 12

```

atgtctggct cctggctgag ggatgtctgg gactggatct gcaactgtgct gactgacttc 60
aagacctggc tgcattccaa gctgctgccc aggtgcctg gcgacccatt cttctcctgc 120
cagaggggct acaggggcgt ctggaggggc gatggcgtga tgcagaccac ctgcccattgt 180
ggcgcccaga tcaactggcca tgtgaagaat ggctccatga ggattgtggg cccaagacc 240
tgctccaaca cctggcatgg caccttcccc atcaatgcct acaccactgg cccatgcacc 300
ccatcccctg cccccaacta ctccagggcc ctgtggaggg tggctgctga ggagtatgtg 360
gaggtgacca ggggtggcga cttccactat gtgactggca tgaccactga caatgtgaag 420
tgcccatgcc aggtgcctgc cctgagttc ttcactgagg tggatggcgt gaggctgcac 480
aggtatgccc ctgcctgcaa gccctgctg agggatgagg tgaccttcca ggtgggcctg 540

```

```

aaccagttcc ctgtgggctc ccagctgcca tgtgagcctg agcctgatgt gactgtgctg 600
acctccatgc tgactgagcc atcccatc atcgtctgaga ctgccaagag gaggctggcc 660
aggggctccc ctccatccct ggctcctccc tctgcctccc agctgtctgc tccatccctg 720
aaggccacct gcaccaccag gcatgactcc cctgatgctg acctgattga ggccaacctg 780
ctgtggaggg aggagatggg cggcaacatc accaggggtg agtctgagaa caaggtgggtg 840
atcctggact cctttgagcc cctgagggct gaggaggatg agagggaggt ctctgtggct 900
gctgagatcc tgaggaagtc caggaagttc cccctgccc tgcccatctg ggcgaggcca 960
tcctacaacc caccctgct ggagtcctgg aaggaccctg actatgtgcc ccctgtggtg 1020
catggctgcc ccctgcccc caccatggcc ccacccatcc cccacccag gaggaagagg 1080
actgtgggtg tgactgagtc cactgtctcc tctgcctgg ctgagctggc caccaagacc 1140
ttcggctcct ctggctcctc tgctgtggac tctggcactg ccacggcccc ccctgaccag 1200
ccatctgatg atggcgacag gggctctgat gatgagtcct actcctccat gccccccctg 1260
gagggcgagc ctggcgaccc tgacctgtct gatggctcct ggtccactgt ctctgaggag 1320
gcctctgagg atgtggcctg ctgctcctaa 1350

```

<210> 13

<211> 449

<212> PRT

<213> Hepatitis C Virus

<400> 13

```

Met Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys Thr Val
 1              5              10              15

Leu Thr Asp Phe Lys Thr Trp Leu His Ser Lys Leu Leu Pro Arg Leu
      20              25              30
Pro Gly Asp Pro Phe Phe Ser Cys Gln Arg Gly Tyr Arg Gly Val Trp
      35              40              45
Arg Gly Asp Gly Val Met Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile
      50              55              60
Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro Lys Thr
      65              70              75              80
Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr
      85              90              95
Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp
      100              105              110
Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly Asp Phe
      115              120              125
His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro Cys Gln
      130              135              140
Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg Leu His
      145              150              155              160
Arg Tyr Ala Pro Ala Cys Lys Pro Leu Leu Arg Asp Glu Val Thr Phe
      165              170              175
Gln Val Gly Leu Asn Gln Phe Pro Val Gly Ser Gln Leu Pro Cys Glu
      180              185              190
Pro Glu Pro Asp Val Thr Val Leu Thr Ser Met Leu Thr Glu Pro Ser
      195              200              205
His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro
      210              215              220
Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu
      225              230              235              240
Lys Ala Thr Cys Thr Thr Arg His Asp Ser Pro Asp Ala Asp Leu Ile
      245              250              255
Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg

```

[illegible]

<210> 14

<211> 1773

<212> DNA

<213> Artificial Sequence

 $\langle 220 \rangle$

<223> Optimized sequence encoding HCV NS5b protein

<400> 14

atgtctctaca	cctggactgg	cgccctgata	accccatgtg	ctgctgagga	gtccaagctg	60
cccatcaacc	cctgtccaa	ctccctgctg	aggcatcaca	acatggtcta	tgccaccacc	120
tccaggtctg	ctggcctgag	gcagaagaag	gtgacctttg	acaggctgca	tgtgcctgat	180
gaccactaca	gggatgtgct	gaaggagatg	aaggccaagg	cctccactgt	gaaggcgaag	240
ctgctgtctg	tggaggaggc	ctgcaagctg	acccctcccc	actctgccag	gtccaagttt	300
ggctatggcg	ccaaggatgt	gaggaacctg	tcttccaagg	ctgtgaacca	catccactct	360
gtctggaag	acctgctgga	ggacactgag	accccattg	acaccaccat	catggccaag	420
aatgaggtct	tctgtgtgca	gcctgagaag	ggcggcagga	agcctgccag	ggtgattgtc	480
tcttcctgac	tgggcgtgag	ggtgtgtgag	aagatggccc	tgtatgatgt	gctctccacc	540
ctgccccagg	ctgtgatggg	ctctctctat	ggcttccagt	actcccctgg	ccagagggtg	600
gagttcctgg	tgaatgcctg	gaagtccaag	aagaacccca	tgggctttgc	ctactgcacc	660
aggtgctttg	actccactgt	gactgagtct	gacatcaggg	tggaggagtc	catctaccag	720
tgctgtgacc	tggctcctga	ggccaggcag	gtgatcaggt	ccctgactga	gaggctgtac	780
attggcgggc	ccctgaccaa	ctccaagggc	cagaactgtg	gctacaggag	gtgcagggcc	840
tctggcgtgc	tgaccactaa	ctgtggcaac	accctgacct	gctacctgaa	ggcctctgct	900
gcttgccagg	ctgccaaagt	gcactgactgc	accatgtctg	tctgtggcga	tgacctggtg	960
gtgatctgtg	agtgctgctg	caccagaggag	gatgtgcct	ccctgagggt	cttacttgag	1020
gccatgacca	ggtactctgc	ccccctggc	gacctcccc	agcctgagta	tgacctggag	1080
ctgatcacct	cctgctcctc	caatgtctct	gtggcccatg	atgcctctgg	caagagggtc	1140
tactacctga	ccagggaccc	caccaccccc	ctggccaggg	ctgcctggga	gactgccagg	1200

```

cacacccctg tgaactcctg gctgggcaac atcatcatgt atgccccac cctgtggggc 1260
aggatgatcc tgatgaccca cttcttctcc atcctgctgg cccaggagca gctggagaag 1320
gccctgggct gccagattta tggcgccacc tacttcattg agccctgga cctgccccag 1380
atcatccaga ggctgcatgg cctgtctgcc ttctccctgc actcctactc ccctggcgag 1440
atcaacaggg tggcctcctg cctgaggaag ctgggcgtgc cccccctgag ggtgtggagg 1500
cacagggcca ggtctgtgag ggccaagctg ctgtcccagg gcggcagggc tgccacctgt 1560
ggcaagtacc tgttcaactg ggctgtgagg accaagctga agctgacccc catccctgt 1620
gcctcccagc tggacctgtc tggctggttt gtggctggct actctggcgg cgacatctac 1680
cactccctgt ccagggccag gccaggtgg ttcatgtggt gcctgctgct gctgtctgtg 1740
ggcgtgggca tctacctgct gcccaacagg tga 1773

```

<210> 15

<211> 590

<212> PRT

<213> Hepatitis C Virus

<400> 15

```

Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu
1      5      10      15
Glu Ser Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Leu Arg His
20     25     30
His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly Leu Arg Gln
35     40     45
Lys Lys Val Thr Phe Asp Arg Leu His Val Pro Asp Asp His Tyr Arg
50     55     60
Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys
65     70     75     80
Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala
85     90     95
Arg Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser
100    105    110
Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu Glu Asp
115    120    125
Thr Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe
130    135    140
Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val
145    150    155    160
Phe Pro Glu Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp
165    170    175
Val Val Ser Thr Leu Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe
180    185    190
Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys
195    200    205
Ser Lys Lys Asn Pro Met Gly Phe Ala Tyr Cys Thr Arg Cys Phe Asp
210    215    220
Ser Thr Val Thr Glu Ser Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln
225    230    235    240
Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Val Ile Arg Ser Leu Thr
245    250    255
Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn
260    265    270
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Asn Cys
275    280    285
Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys Arg Ala
290    295    300

```

```

Ala Lys Leu His Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val
305          310          315          320
Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu Arg
          325          330          335
Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro
          340          345          350
Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn
          355          360          365
Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr
          370          375          380
Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg
385          390          395          400
His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro
          405          410          415
Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Ile Leu
          420          425          430
Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Gly Cys Gln Ile Tyr Gly
          435          440          445
Ala Thr Tyr Phe Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg
          450          455          460
Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu
465          470          475          480
Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu
          485          490          495
Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Lys Leu Leu Ser
          500          505          510
Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala
          515          520          525
Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu
          530          535          540
Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr
545          550          555          560
His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu
          565          570          575
Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
          580          585          590

```

<210> 16
 <211> 103
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified Vector Sequence

<400> 16
 cttaagcgaa gggtactctt gtacctctgg tacttggtcg ggatggtgta gacggcgccg 60
 aagtggacga agttcttcat ttgggccctt aagatttcag ctg 103

<210> 17
 <211> 15
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 17

agatctacca tgagc

15

<210> 18

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 18

gccgaattcg cttcc

15

<210> 19

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 19

taaaccggg aattctaaag tcgac

25

<210> 20

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 20

atcaccatgg at

12

<210> 21

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 21

gagatcttca tgagc

15

<210> 22

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified Vector Sequence

<400> 22
agatccacca tgcag

15

<210> 23
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified Vector Sequence

<400> 23
ggtgcagatc tgatgagc

18

<210> 24
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified Vector Sequence

<400> 24
gcctaaagtc gac

13

<210> 25
<211> 4261
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified Vector Sequence

<400> 25
gatatgggct attggccatt gcatacgttg tatccatata ataatatgta catttatatt 60
ggctcatgtc caacattacc gccatgttga cattgattat tgactagtta ttaatagtaa 120
tcaattacgg ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg 180
gtaaatggcc cgcctggctg accgcccaac gacccccgcc cattgacgtc aataatgacg 240
tatgttccca tagtaacgcc aatagggaact ttccattgac gtcaatgggt ggagtattta 300
cggtaaaactg cccacttggc agtacatcaa gtgtatcata tgccaagtac gccccctatt 360
gacgtcaatg acggtaaatg gcccgccctgg cattatgccc agtacatgac cttatgggac 420
tttctactt ggcagtacat ctacgtatta gtcacgcta ttaccatggg gatgcgggtt 480
tggcagtaca tcaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccac 540
cccattgacg tcaatgggag tttgttttgg cacaaaaatc aacgggactt tccaaaatgt 600
cgtaacaact cgcgcccatg gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat 660
ataagcagag ctgctttagt gaaccgtcag atcgccctgga gacgccatcc acgctgtttt 720
gacctccata gaagacaccg ggaccgatcc agcctccgcg gccgggaacg gtgcattgga 780
acgcggatcc cccgtgccaa gagtgacgta agtaccgcct atagagtcta taggcccacc 840
cccttggctt cttatgcatg ctatactgtt tttggcttgg ggtctatata cccccgcttc 900
ctcatgttat aggtgatggg atagcttagc ctataggtgt gggttattga ccattattga 960
ccactccct attgggtgacg atactttcca ttactaatcc ataacatggc tctttgccac 1020
aactctcttt attggctata tgccaatata ctgtccttca gagactgaca cggactctgt 1080
atttttacag gatgggggtct catttattat ttacaaattc acatatacaa caccaccgtc 1140

```

cccagtgcc  gcagttttta  ttaaacataa  cgtgggatct  ccacgcgaat  ctcggggtacg  1200
tgttccggac  atgggctctt  ctccggtagc  ggcggagctt  ctacatccga  gccctgctcc  1260
catgcctcca  ggcactcatg  gtcgctcggc  agctccttgc  tcctaacagt  ggaggccaga  1320
cttaggcaca  gcacgatgcc  caccaccacc  agtgtgccgc  acaaggccgt  ggcggtagg  1380
tatgtgtctg  aaaatgagct  cggggagcgg  gcttgcaccg  ctgacgcatt  tggaagactt  1440
aaggcagcgg  cagaagaaga  tgcaggcagc  tgagttgttg  tgttctgata  agagtcagag  1500
gtaactccc  ttgcggtgct  gttaacgggt  gagggcagtg  tagtctgagc  agtactcgtt  1560
gctgccgcgc  gcgccaccag  acataatagc  tgacagacta  acagactgtt  cctttccatg  1620
ggtcttttct  gcagtcaccg  tccttagatc  taccatgagc  accaacccca  agccccagag  1680
gaagaccaag  aggaacacca  acaggaggcc  ccaggatgtg  aagttccctg  ggggaggcca  1740
gattgtggga  ggggtctacc  tgctgccag  gaggggcccc  aggctggggg  tgagggtac  1800
caggaagacc  tctgagaggt  cccagcccag  gggcaggagg  cagcccatcc  ccaaggccag  1860
gaggcctgag  ggccgctcct  gggcccagcc  tggctacccc  tggcccctgt  atggcaatga  1920
aggctttggc  tgggctggct  ggctgctgtc  ccccaggggc  tccaggccct  cctggggccc  1980
cacagacccc  aggaggaggt  ccaggaacct  gggcaagggt  attgacaccc  tgacctgtgg  2040
ctttgctgac  ctgatgggct  acatccccct  ggtgggggct  cctgtgggag  ggggtggctag  2100
ggctctggct  catgggggtga  ggggtgctga  ggatgggggt  aactatgcta  ctggcaacct  2160
gcttggtgc  tcttcttcca  tcttctgtc  ggcctgtctc  tcctgcctga  cagtgcctgc  2220
ttctgccgaa  ttcgcttcca  atgagaacat  ggagaccatg  aaccagccct  accacatctg  2280
ccgcggcttc  acctgcttca  agaagtaaac  ccgggaattc  taaagtcgac  agcggcccg  2340
atctgctgtg  ccttctagtt  gccagccatc  tgttgtttgc  ccctcccccg  tgccttccct  2400
gacctggaa  ggtgccactc  ccaactgtct  ttcctaataa  aatgaggaaa  ttgcatcgca  2460
ttgtctgagt  aggtgtcatt  ctattctggg  ggggtggggt  gggcagcaca  gcaaggggga  2520
ggattgggaa  gacaatagca  ggcattgctg  ggatgcgggt  ggctctatgg  gtacggccgc  2580
agcggcetta  attaaggccg  cagcggccgt  acccaggtgc  tgaagaattg  acccggttcc  2640
tcgaccgcta  aaaaggccgc  gttgctggcg  tttttccata  ggctccgccc  cctgacgag  2700
catcacaaaa  atcgacgctc  aagtcaagg  tggcgaaacc  cgacaggact  ataaagatac  2760
caggcgtttc  cccctggaag  ctccctcggt  cgctctcctg  ttccgacct  gccgttacc  2820
ggatacctgt  ccgcctttct  cccttcggga  agcgtggcgc  tttctcaatg  ctacgctgt  2880
aggtatctca  gttcggtgta  ggtcgttcgc  tccaagctgg  gctgtgtgca  cgaaccccc  2940
gttcagcccc  accgctgcgc  cttatccggt  aactatcgtc  ttgagtccta  cccggttaaga  3000
cacgacttat  cgccactggc  agcagccact  ggtaacagga  ttagcagagc  gaggtatgta  3060
ggcggtgcta  cagagttctt  gaagtgggtg  cctaactacg  gctacactag  aaggacagta  3120
tttggtatct  gcgctctgct  gaagccagtt  accttcggaa  aaagagttgg  tagctcttga  3180
tccggcaaaa  aaaccaccgc  tggtagcgg  ggtttttttg  tttgcaagca  gcagattacg  3240
cgcagaaaaa  aaggatctca  agaagatcct  ttgatctttt  ctacgtgatc  ccgtaatgct  3300
ctgccagtgt  tacaaccaat  taaccaattc  tgattagaaa  aactcatcga  gcatcaaattg  3360
aaactgcaat  ttattcatat  caggattatc  aataccatat  ttttgaaaaa  gccgtttctg  3420
taatgaagga  gaaaactcac  cgaggcagtt  ccataggatg  gcaagatcct  ggtatcggtc  3480
tgcgattccg  actcgtccaa  catcaatata  acctattaat  ttccctcgt  caaaaataag  3540
gttatcaagt  gagaaatcac  catgagtga  gactgaatcc  ggtgagaatg  gcaaaaagctt  3600
atgcatttct  ttccagactt  gttcaacagg  ccagccatta  cgctcgtcct  caaaatcact  3660
cgcatacaacc  aaaccgttat  tcattcgtga  ttgcgcctga  gcgagacgaa  atacgcgatc  3720
gctgttaaaa  ggacaattac  aaacaggaat  cgaatgcaac  cggcgcagga  aactgcccag  3780
cgcatacaaca  atattttcac  ctgaatcagg  atattcttct  aatacctgga  atgctgtttt  3840
cccgggggatc  gcagtgggtga  gtaaccatgc  atcatcagga  gtacggataa  aatgcttgat  3900
ggtcggaaga  ggcataaatt  ccgtcagcca  gtttagtctg  accatctcat  ctgtaacatc  3960
attggcaacg  ctacctttgc  catgtttcag  aaacaactct  ggcgcatcgg  gcttcccata  4020
caatcgatag  attgtcgcac  ctgattgccc  gacattatcg  cgagcccat  tatacccata  4080
taaatacagca  tccatgttgg  aatttaatcg  cggcctcgag  caagacgttt  cccgttgaat  4140
atggctcata  acacccttg  tattactgtt  tatgtaagca  gacagtttta  ttgttcatga  4200
tgatatattt  ttatcttgtg  caatgtaaca  tcagagattt  tgagacacaa  cgtggctttc  4261

```

c